DOI: 10.63550/ICEIP.2025.63.41.005

Assessment of Digitalization of Paid Services and Services of the Regions of the Russian Federation Based on a Multiparametric Data Analysis

Lyudmila Borisova[a], Irina Sedykh[b], Marina Khripunova[c]
Financial University under the Government of the Russian Federation, 49 Leningradsky Ave., Moscow, 125993, Russia

Keywords: statistics; mathematical model; machine learning methods; evaluation of parameters.

Abstract: Currently, there is a transition from analogue to digital technologies in all spheres of life. The pace of life is accelerating, and more and more digital technologies are found in key areas of life, such as education and healthcare. Therefore, it is important to look at how things are with the digitalization of knowledge-intensive spheres of life, such as science, education, healthcare, and telecommunications technologies in the context of the regions of the Russian Federation. The paper uses machine learning methods to classify regions according to a set of indicators of electronic services and services. A clear division of the regions into two large, almost equal groups according to this set of indicators has been obtained. The use of well-known statistical criteria has demonstrated the statistical significance of such a division. Scattering diagrams are constructed as an example of the relationship of such indicators. The multiple correlation coefficient between the indicators of electronic services and services is 0.71, which indicates a close relationship between the indicators of digitalization of services. In addition, the division of regions into clusters was obtained using hierarchical clustering, which implies that Moscow has significantly overtaken other regions of Russia in providing electronic services and services, and the remaining regions are heterogeneous in this indicator, taking into account the considered indicators of digitalization of regions according to Rosstat data for July 2024.

1 INTRODUCTION

The main objectives of the implementation of the concept of regional informatization are to improve the quality of life of citizens through the use of information technology. Thanks to the introduction of innovative technologies into informatization, a number of macroeconomic problems can be solved: unemployment, inflation, economic growth and the national product. The authors of this work proposed a mechanism for informatization of innovative technologies to ensure macroeconomic trends in a closed-loop economy, which includes several stages of determining their content, opportunities for optimization in accordance with the needs and pace of macroeconomic trends.

A characteristic feature of the modern stage of society's development is the digitalization of the economic and social spheres. Digital transformation has become an important factor in global economic growth: the potential economic effects of the digital economy can significantly increase GDP, people's purchasing power, change the market and quality of life, as well as improve the business environment. The emergence of the concept of “digital economy” marked a new stage in the management of production of goods and services based on the use of modern information technologies.

The introduction of digital technologies allows you to gain competitive advantages. In Russia, the leaders in the use of digital technologies are primarily financial institutions. Sberbank and T-bank are the leaders. When using IT technologies, changes are taking place in the fields of healthcare and education. Digitalization of social services includes not only digital technologies, but also people themselves. Currently, there are more and more studies devoted to the successful application of IT technologies in

[a] https://orcid.org/0000-0002-5757-0341
[b] https://orcid.org/0000-0002-5912-2614
[c] https://orcid.org/0009-0006-9848-7495

medicine. For example, (Davis, Roudsari, Raworth, Courtney, MacKay, 2017) proposes an innovative digital platform that defines patient interaction and management of digital medical records in a new way. The use of computer science in healthcare helps to improve the quality of services provided, find new diagnostic methods and improve methods of care and treatment. Despite the numerous advantages of such integration, it also has many disadvantages listed in (Chinnakali, Kumar, 2024).

The article (Zhang, van Gorp, Kievit, 2022) shows that the level of digital technology development is positively related to the performance of national entrepreneurial ecosystems, and this positive relationship is strengthened in countries with a favorable culture, high-quality institutions, supportive policies, affordable resources and a well-developed service industry. This conclusion was made based on the analysis of socio-economic data for 101 countries for the period from 2001 to 2018. The paper considers the correlation coefficients between the studied indicators and checks the significance of these coefficients.

Regression analysis was applied in (Toth, Elekes, Whittle, Lee, 2022). The results of the regression analysis indicate that Italian regions with a more developed network of technological knowledge demonstrate a higher level of resilience to changes in employment levels.

A high quality of human life also includes access to knowledge, education and cultural values that shape personality and ideas about the world around us. This is influenced by the amount of funds allocated for education: the larger it is, the more advanced technologies are introduced into education, the faster students will be able to adapt to modern society.

This paper presents the results of recognition of the regions of the Russian Federation by a set of indicators of electronic services and services for July 2024.

2 MATERIALS AND METHODS

Materials. The open data of Rosstat on indicators of the socio-economic development in 86 regions of the Russian Federation are used for the study. We used 8 indicators, namely the following paid services in calculating the number of rubles per capita: educational, telecommunications, postal services and couriers, physical education and sports, sports clubs and fitness centers, medical, veterinary and legal. Educational services were taken as a grouping indicator. All 86 regions of the Russian Federation were divided into two groups: the first group with the indicator of paid services per capita below the median level (291 rubles), and the second group above this value. There were 44 regions in the first group, 42 in the second.

Methods. The method of statistically weighted syndromes (SHS) was used (Kuznetsova, Senko, 2005; Kuznetsova, Senko, Kuznetsova, 2015; Kuznetsova, Borisova, Kuznetsova, Senko, 2017; Senko, Kuznetsova, 2010).

3 RESULTS

3.1 Mann-Whitney-Wilcoxon statistical test

Note that at first we considered 10 indicators. In addition to these eight, hotel services and similar services for the provision of temporary housing, paid services for the elderly and disabled were analyzed, but these indicators turned out to be statistically insignificant (Table 1.)

Table 1: Mann-Whitney-Wilcoxon test.

Indicator	Mean in group 1	Standard deviation in group 1	Mean in group 2	Standard deviation in group 2	p-value
Medicalal services, rubles, capita	498,9	183,3	796,6	385,3	0,00001
Telecommunicational services, rubles, capita	791,1	268,5	949,8	328	0,01
Postal services and couriers services, rubles, capita	42,1	35,5	50,2	38,9	0,06
Physical education and sports services, rubles, capita	64,17	105,1	78,2	57,5	0,005
Sports clubs and fitness centers services, rubles, capita	23,9	22,3	40,4	38,9	0,007
Veterinary services, rubles, capita	20,17	11,14	28,8	12,8	0,002
Legal services, rubles, capita	62,72	21,3	87,92	43,2	0,006
Hotel services and similar temporary housing services, rubles, capita	243,5	451,1	321,9	509,3	0,38
Paid services for the elderly and the disabled, rubles, capita	39,5	23,6	38,3	33,0	0,24

Note that the significant coefficient of multiple regression between the grouping indicator and the indicators indicated in Table 1 is 0.71, which indicates a close relationship between the studied features (p-value=0).

The recognition quality is usually calculated using the area under the ROCK curve (in relative units). The closer this value is to 1, the better the recognition. Figure 1 shows the ROCK curve obtained using the SHS method.

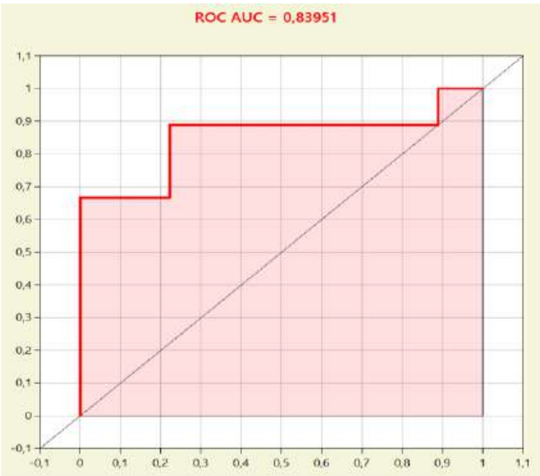

Figure 1: ROC-curve.

The Voronezh Region was incorrectly recognized. In July, expenses for paid educational services were above the median level. By mistake, she was assigned to the 1st class.

Note that the sensitivity of recognition of regions belonging to class 1 (with a grouping attribute value below the median) was significantly worse.

Kaluga, Smolensk, Oryol, Kirov, Irkutsk Regions, Kamchatka Territory, the Republics of Yakutia (Sakha) and Buryatia were incorrectly recognized. Paid education services in July 2024 in these regions were below the median level, and the costs of other paid services were high.

In this regard, a cluster analysis was performed, the results of which are shown in Fig.2.

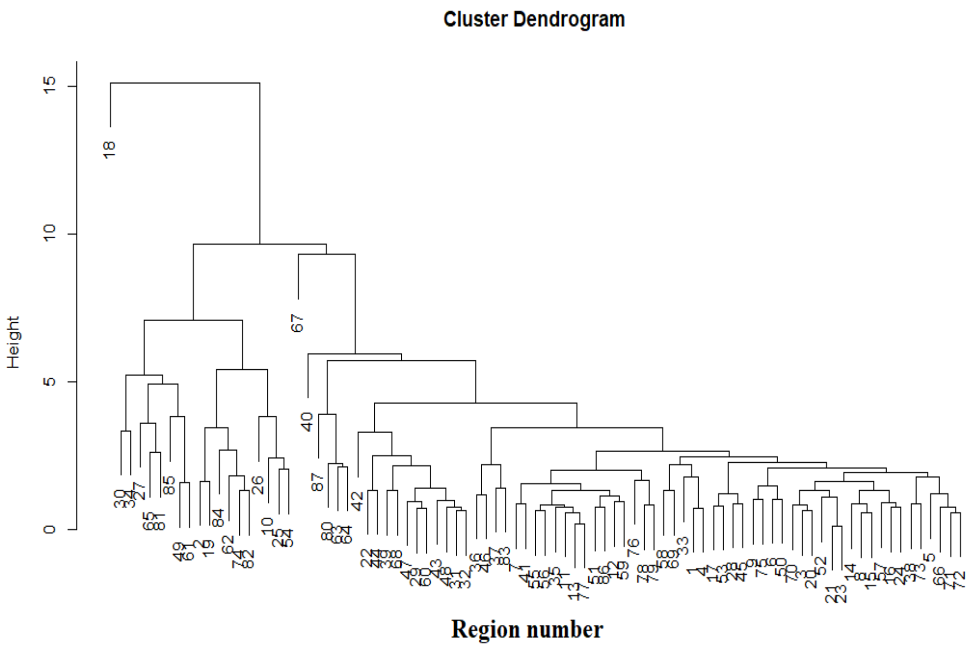

Figure 2: Cluster dendrogram.

As follows from the dendrogram. only Moscow differs significantly from all regions of Russia (number 18 on the dendrogram).

It is interesting to look at the scattering diagrams to imagine the impact of the studied indicators of digitalization on each other.

Figure 3 shows a scattering diagram for two features. One-dimensional scattering diagrams were constructed for other pairs of studied features. They also indicated the presence of a threshold value that must be overcome for the transition of the region from one class to another (the results of these calculations are not presented in the article).

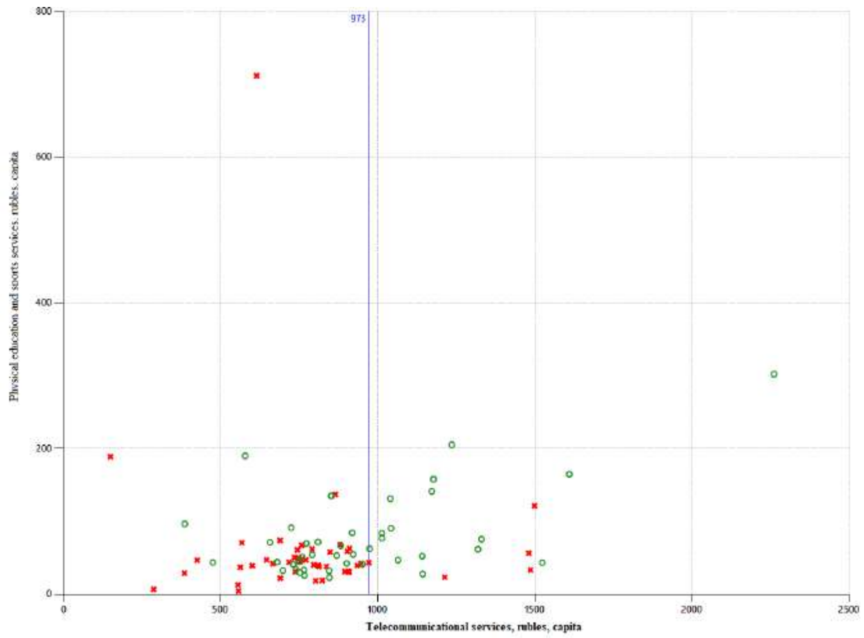

Figure 3: Scattering diagram.The first class prevails in the first quadrant. The second quadrant is dominated by the second class.

The scattering diagram (Fig. 3) clearly shows the values for the regions of the first group (red crosses) and the second (green circles) according to the corresponding indicators. The boundaries of the division are set in such a way that objects of the same class predominate on one side of the border. To transfer a region from the first group to the second, it is necessary to increase the population's spending on telecommunications services (they should exceed the value of 973 rubles per capita per month).

4 CONCLUSIONS

Our research has shown that regions as a whole are recognized by indicators of digital services among the regions of the Russian Federation. The recognition quality indicator ROC AUC=0.84.

The main objectives of the implementation of regional informatization are to improve the quality of life of citizens through the use of information and technology, as well as to equalize the level of development of the information society in the subjects of the Russian Federation.

The dendrogram obtained using cluster analysis shows a significant difference in the level of digital services and services in Moscow compared to other regions of the Russian Federation, which do not differ significantly from each other in the eight indicators studied. with the exception of the Altai Territory (region 67 on the dendrogram), perhaps due to the presence of a large number of small villages and villages that have not been affected by digitalization to the same extent as for other regions.

Machine learning alone is not enough to classify regions by the level of digitalization in the field of electronic services and services. The results obtained using the method of statistically weighted syndromes, despite the high value of the ROCAUC recognition quality index, demonstrated a large number of errors when using sliding control, when 20% of regions were randomly selected for pure control. Significant errors were found only when recognizing regions belonging to the first group. Perhaps it so happened that in the summer, during the holidays, the number of people using paid educational services. It was significantly less than in autumn-winter-spring. A comparative analysis was carried out using the above indicators of digital services for May 2024. It turned out that the sensitivity of the method of statistically weighted syndromes for this period was higher, since only three recognition errors were obtained.

It seems useful to find multiple regression coefficients, especially since the analysis of foreign and domestic scientific literature has shown the widespread use of regression models in studying the quality of life of the population in various regions. It should be noted that the obtained high coefficient of multiple correlation between the indicator of digitalization of educational services and eight indicators is statistically significant and demonstrates the presence of a close relationship between the studied features.

It is also useful to use cluster analysis to visualize the differences between regions. However, cluster analysis alone cannot take into account the presence of hidden relationships that can be identified using machine learning methods.

Scattering diagrams are also of interest, thanks to which you can clearly see in which direction one or another indicator in the field of digital services and services should be changed in order for the region to move from one class to another.

The results of the conducted research can be used by state executive authorities in the process of updating state programs for the development of specific regions in the direction of the development of the digital economy.

REFERENCES

Kuzior, A., Arefiev, S., Poberezhna, Z., 2023. Informatization of innovative technologies for ensuring macroeconomic trends in the conditions of a circular economy. *Journal of open innovation: technology, market and complexity.* V. 9. N.1. 100001.

Uskov, V.S., 2022. Development of the information society in the Russian Federation: problems and prospects. *Economic and social changes: facts, trends, forecast.* V. 15. N. 2. P. 120-137.

Petrova, N. P., Ovechkina, A. I., 2021. On the issue of the digital transformation of the Russian economy. *Proceedings of the St. Petersburg State University of Economics.* No. 1. P. 57-62.

Romanova, N. V., 2020. Digitalization of social services: problems and prospects. *Bulletin of the USPTU. Science, education, economics. Economics series.* No. 1 (31). pp. 58-65.

Davis, S., Roudsari, A., Raworth, R., Courtney, K.L., MacKay, I., 2017. Shared decision-making using personal health record technology: a scoping review at the crossroads. *J Am Med Inform Assoc.* Jul 01. 24(4).857-866.

Chinnakali, P., Kumar, S.S., 2024. Principles and Application of Evidence-based Public Health Practice. 2024. Academic Press. Edited by S. Kathirvel, A. Singh and A. Chockalingam. Chapter 14 - Effective use of information technology for the quality of public health practice/ P/ 239-259.

Zhang, J, van Gorp, D., Kievit, H., 2022. Digital technology and national entrepreneurship: An ecosystem perspective. *The Journal of Technology Transfer.* 48(3):1-29.

Toth, G., Elekes, Z., Whittle, A., Lee, C., 2022. Technology Network Structure Conditions the Economic Resilience of Regions. *Economic Geography.* 98(2).

Zholudeva, V.V., Melnichenko, n.F., Kozlov, G.E., 2015. Statistical estimation of the life quality of central federal district population. *Economics, Statistics and Computer Science.* Vol.173. No. 2.

Kuznetsova, A.V., Senko, O.V., 2005. Possibilities of using Data Mining methods in medical laboratory research to identify patterns in data arrays. *The doctor and information technologies.* Vol. 1. pp. 49-56.

Kuznetsova, A.V., Senko, O.V., Kuznetsova, Yu.O., 2015. Overcoming the black box problem when using machine learning methods in medicine. *The journal "Doctor and information technologies. Artificial intelligence in healthcare".* No. 7 (October). Pp.74-80.

Kuznetsova, Yu.O., Borisova, L.R., Kuznetsova, A.V., Senko, O.V., 2017. Transparent interface for forecasting in machine learning. Moscow State University. Collection of scientific papers of the XIX International Conference DAMDID/RCDL' pp. 493-495.

Senko, O., Kuznetsova, A. A., 2010. Recognition method based on collective decision making using systems of regularities of various types. *Pattern Recognition and Image Analysis.* V. 20. № 2. P. 152-162.